

# 基于麦克风阵列的 GCC 时延估计算法分析<sup>①</sup>



唐浩洋, 陈子为, 黄 维

(成都信息工程大学 电子工程学院, 成都 610225)

通讯作者: 陈子为, E-mail: czw@cuit.edu.cn

**摘 要:** 准确的时延估计 (Time Delay Estimation, TDE) 是基于到达时间差 (Time Difference of Arrival, TDOA) 的声源定位技术的前提。在众多时延估计算法中, 广义互相关 (Generalized Cross Correlation, GCC) 算法因其较低的运算复杂度和易于实现的特点得到了广泛的应用。针对不同的噪声情况, GCC 时延估计算法利用不同的加权函数来抑制噪声干扰。本文在介绍麦克风阵列模型和 GCC 时延估计算法的基础上, 针对 GCC 算法的弊端提出了一种改进算法, 并在多种信噪比条件下, 对部分加权函数的 GCC 时延估计算法进行了 MATLAB 仿真, 通过比较其时延估计性能和声源定位精度, 分析了这些加权函数各自的优劣性。

**关键词:** 麦克风阵列; 声源定位; 时延估计; 广义互相关; 加权函数

引用格式: 唐浩洋, 陈子为, 黄维. 基于麦克风阵列的 GCC 时延估计算法分析. 计算机系统应用, 2019, 28(12): 140-145. <http://www.c-s-a.org.cn/1003-3254/7173.html>

## Analysis of GCC Time Delay Estimation Algorithm Based on Microphone Array

TANG Hao-Yang, CHEN Zi-Wei, HUANG Wei

(College of Electronic Engineering, Chengdu University of Information Technology, Chengdu 610225, China)

**Abstract:** Accurate Time Delay Estimation (TDE) is a prerequisite for sound source localization technology based on Time Difference Of Arrival (TDOA). Compared with other TDE algorithms, the Generalized Cross Correlation (GCC) algorithm has been widely used due to its low computational complexity and implementation simplicity. GCC time delay estimation algorithm uses different weighting functions to suppress noise and other interference in different noise conditions. This work first presents microphone array model and GCC algorithm. Then, an improved algorithm proposed in the study is elaborated in view of the disadvantages of GCC algorithm. Furthermore, GCC algorithm to partial weighting functions is simulated using MATLAB under different Signal-to-Noise Ratio (SNR) conditions. The advantages and disadvantages of these weighting functions are analyzed by comparing the performance of time delay estimation and the accuracy of sound source localization.

**Key words:** microphone array; sound source localization; Time Delay Estimation (TDE); Generalized Cross Correlation (GCC); weighting functions

基于麦克风阵列的声源定位技术在人工智能、雷达探测、电话会议等多个领域有广泛的应用。根据 Brandstein 的分类方法<sup>[1]</sup>, 将基于麦克风阵列的声源定位技术分为 3 类: 基于到达时间差的定位技术 (TDOA)、

基于最大输出功率的可控波束形成技术和基于高分辨率谱估计的定位技术。基于到达时间差的定位方法在定位准确性上, 优于另外 2 种方法<sup>[2]</sup>, 并且由于其计算量相对较小、在实时系统中可以实现等原因, 在声源

① 基金项目: 四川省教育厅科学研究重点项目 (18ZA0088)

Foundation item: Major Project of Science and Technology Research Program of Education Bureau, Sichuan Province (18ZA0088)

收稿时间: 2019-04-13; 修改时间: 2019-05-16, 2019-05-27; 采用时间: 2019-05-30; csa 在线出版时间: 2019-12-10

定位方面受到了更多的青睐,被广泛使用.对于 TDOA,获得准确的时延估计是声源定位精确的前提条件.根据所采用的不同物理参量可将时延估计分为 2 类: 1) 基于广义互相关函数 (Generalized Cross Correlation, GCC)<sup>[3]</sup>的时延估计算法; 2) 基于自适应最小均方 (Least Mean Square, LMS)<sup>[4]</sup>的时延估计算法.在实际应用中,由于噪声的干扰和语音信号的短时处理,可能会使 GCC 时延估计算法出现较大的误差,通常的做法是引入加权函数来抑制噪声干扰.本文介绍了麦克风阵列时延估计模型和广义互相关时延估计算法以及一种改进算法,并对这几种加权方法进行了 Matlab 仿真,通过比较各加权方法的时延估计效果以及声源定位精度,分析其优劣性.

## 1 基于麦克风阵列的声源定位算法

麦克风阵列中,当不同阵元接收同一声源信号时,由于信号传输距离的不同引起了时间上的差别.通过声源信号到达各麦克风的时间差(即时延)和声音在媒质中的传播速度即可估计声源的方向和位置,从而实现声源定位.在这个过程中准确的时延估计是基于 TDOA 的声源定位技术的前提.

### 1.1 麦克风阵列声源定位模型

麦克风阵列是由麦克风作为阵元按一定的方法排列组成的阵列.根据声源离麦克风的距离远近,可将声源传递的波分为近场波和远场波.假设声源距麦克风阵列较远,产生的是远场波.此时可以将远场波近似看作成平面波.此外,为了简化分析,将麦克风阵列抽象为均匀线性阵列 (Uniform Linear Array, ULA),该阵列由同一直线上距离相同间隔的多个麦克风组成.构建的麦克风阵列模型如图 1 所示<sup>[2]</sup>.

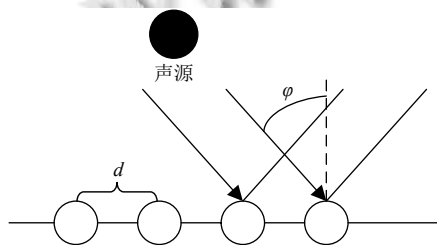


图 1 麦克风阵列模型

图 1 中,  $d$  表示两阵元的间距,一般通过测量得到,是已知值;  $\varphi$  表示声源信号与麦克风阵列法线的夹角.

由图 1 可知,对于两个相邻的阵元,语音信号在传

播过程中存在  $d \sin \varphi$  的距离差,这也就使得两相邻阵元在接收信号时存在一个时间延迟,即:

$$\tau(\varphi) = \frac{d \sin \varphi}{c} \quad (1)$$

式中,  $c$  表示声音的传播速度,此处假设声音在 1 个标准大气压和 15 °C 的条件下传播,即  $c=340$  m/s.由式 (1) 可知,通过计算时间延迟即可估计声源的方向和位置.

### 1.2 基于互相关的时延估计算法

在基于 TDOA 的声源定位算法中,Knapp 和 Carter<sup>[3]</sup>提出的广义互相关函数的时延估计算法是一种传统的 TDE 算法.因为来自同一声源的信号必然存在一定的相关性,该算法根据这一特点,通过计算不同阵元接收到的声源信号之间相关函数,利用数学中相关函数的性质,选取使互相关函数最大的时刻,即可估算出不同阵元间的时延差值.

在一个麦克风阵列系统中,任意选取阵列中的 2 个麦克风,假设这 2 个麦克风接收到的信号分别为  $x(n)$  和  $y(n)$ ,则这两路信号可表示为:

$$x(n) = \alpha_x s(n) + v_x(n) \quad (2)$$

$$y(n) = \alpha_y s(n - T) + v_y(n) \quad (3)$$

其中,  $n$  表示接收到的数字信号的时间采样点;  $s(n)$  表示阵列接收到的有用信号;  $v_x(n)$  和  $v_y(n)$  表示两个阵元在接收时产生的加性噪声,此处将噪声假定为零均值平稳噪声且与有用信号  $s(n)$  相互独立;  $\alpha_x$  和  $\alpha_y$  表示声音在传播过程中的衰减系数 ( $\alpha_x < 1, \alpha_y < 1$ ),  $T$  表示两个阵元接收到的有用信号之间的时延.

对于两个麦克风阵元,其接收到的信号  $x(n)$  和  $y(n)$  的互相关函数  $R_{xy}(\tau)$  可表示为:

$$R_{xy}(\tau) = E[x(n)y(n - \tau)] \quad (4)$$

其中,  $E[\cdot]$  为数学期望.假设有用信号  $s(n)$  和加性噪声  $v(n)$  相互独立,在无混响条件下,将式 (2) 和式 (3) 代入到式 (4) 中,可得:

$$\begin{aligned} R_{xy}(\tau) &= \alpha_x \alpha_y E[s(n)s(n - T - \tau)] + E[v_x(n)v_y(n - \tau)] \\ &= \alpha_x \alpha_y R_{ss}(\tau - T) + R_{v_x v_y}(\tau) \end{aligned} \quad (5)$$

其中,  $R_{ss}(\tau - T)$  表示声源的有用信号  $s(n)$  的自相关函数;  $R_{v_x v_y}(\tau)$  表示噪声  $v_x(n)$  和  $v_y(n)$  的互相关函数.

由式 (5) 可得,若阵元接收到的信号的信噪比足够的大,由相关函数的性质可知,当  $\tau = T$  时,  $R_{ss}(\tau - T)$  取得最大值.所以,求出当  $R_{xy}(\tau)$  取峰值时对应的  $\tau$  就可得到了信号  $x(n)$  和  $y(n)$  之间的时延  $T$ ,即进行了时延估计.

## 2 广义互相关时延估计算法

### 2.1 GCC 时延估计算法

根据维纳-辛钦定理<sup>[5,6]</sup>, 互相关函数及其互功率谱之间的关系可表示为:

$$R_{xy}(\tau) = \int_{-\infty}^{+\infty} G_{xy}(\omega)e^{j\tau\omega} d\omega \quad (6)$$

由于混响、噪声和语音信号的短时处理, 信号中可能会包含多个回波分量. 在这种情况下, 若采用基于互相关函数的时延估计算法计算得到的互相关函数是包含直达波与反射波共同形成的峰值, 这些峰在低信噪比的情况下会降低时延估计的精度. 此外, 对于宽带语音信号, 基音周期的存在使得互相关函数更加复杂<sup>[7]</sup>. 为了保证时延估计不受信号本身的影响, 尽可能地抑制混响等影响, 提高时延估计精度, 需要对该算法进行改进. 常用的方法就是在信号的功率谱中进行加权, 从而在时域中抑制噪声干扰, 锐化相关的信号, 使时延估计准确. 这种方法得到的互相关函数就是广义互相关

函数, 即<sup>[2]</sup>:

$$R_{xy}^{GCC}(\tau) = \int_{-\infty}^{+\infty} \psi_{xy}(\omega)G_{xy}(\omega)e^{j\tau\omega} d\omega \quad (7)$$

其中,  $\psi_{xy}(\omega)$ 表示 GCC 的加权函数. GCC 的基本原理如图 2 所示.

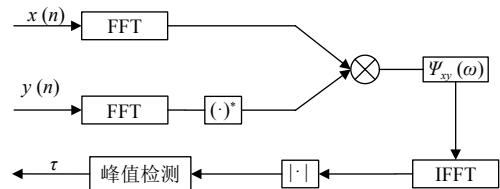


图 2 GCC 基本原理

图 2 中,  $(\cdot)^*$ 表示取共轭运算;  $|\cdot|$ 表示求模运算.

在实际应用中, 针对不同的噪声情况可以选择不同的加权函数 $\psi_{xy}(\omega)$ , 来达到有效抑制噪声的目的. 将几种常用的加权函数以表格的形式进行归纳, 如表 1 所示.

表 1 常用 GCC 加权函数及其特性

名称	$\psi_{xy}(\omega)$	特性
CC	1	对外围噪声、反射和有限观测数据敏感
Roth	$\frac{1}{G_{xx}(\omega)} \text{ 或 } \frac{1}{G_{yy}(\omega)}$	等价于 Wiener 滤波, 可有效抑制噪声大的频段, 但会展开相关函数的峰
SCOT	$\frac{1}{\sqrt{G_{xx}(\omega)G_{yy}(\omega)}}$	相比 Roth, 同时考虑两个通道的影响, 但也会展开相关函数的峰
PHAT	$\frac{1}{ G_{xy}(\omega) }$	相当于白化滤波器, 能够有效抑制噪声, 当信号能量较小时, 分母趋于零, 导致误差增加
ML/HT	$\frac{ \Gamma_{xy}(\omega) }{ G_{xy}(\omega) (1- \Gamma_{xy}(\omega) ^2)}, \Gamma_{xy}(\omega) = \frac{G_{xy}(\omega)}{\sqrt{G_{xx}(\omega)G_{yy}(\omega)}}$	对高信噪比频段赋予大权值, 低信噪比频段赋予小权值, 从而较好的抑制噪声的影响, 但如果有混响存在, 会极大降低时延估计的精度 <sup>[2,8]</sup>
HB	$\frac{ G_{xy}(\omega) }{G_{xx}(\omega)G_{yy}(\omega)}$	对信号周期成分有抑制效果, 低信噪比时效果与 CC 接近

### 2.2 GCC 时延估计流程

语音信号是明显的非平稳信号, 但在 10~30 ms 的时段上, 语音信号是具有短时平稳性的. 因此, 对语音信号进行分帧处理<sup>[9]</sup>是至关重要的. 选取帧长度为 256 个点, 在 8 KHz 采样频率下, 这样大概是 32 ms 长度的信号. 为消除每帧信号的边缘效应, 采用哈密窗 (Hamming) 对输入信号加窗. 时延估计算法流程如图 3 所示.

法复杂度低等优点, PHAT-GCC 时延估计算法被广泛地应用于实际生活. 然而, 该算法仍存在一定的弊端, 当环境噪声过大, 出现混响, 或者信噪比突变的时候, PHAT-GCC 算法的性能也会急剧下降. 基于此, 本文提出了一种改进的 PHAT-GCC 算法.



图 3 时延估计算法流程图

## 3 改进的 PHAT-GCC 时延估计算法

在众多加权函数的广义互相关时延估计算法中, PHAT-GCC 能够更好地抑制噪声, 使互相关函数得到更好的锐化效果, 从而提高时延估计的精度. 加之其算

### 3.1 语音信号预处理

当语音信号受到噪声的干扰, 其最明显的影响就是语音信号的幅值会发生一定的变化. 因为噪声具有



高频率,低幅值的特点,当语音信号某时刻幅值较小,噪声甚至会淹没该时刻信号.所以提出将接收到的语音信号乘上3次方,即 $x(n)$ 转化为 $x^3(n)$ .处理之后,原语音信号中幅值较大的信号更大,而幅值较低的更小,能够有效地抑制噪声.

### 3.2 功率谱的噪声削弱

在实际情况下,尤其是在封闭环境中,噪声会利用墙壁等实体反射,进而导致两个麦克风阵元接收到的噪声并非不相关.因此,在PHAT-GCC算法中,需要把PHAT加权函数中的互功率谱减去相关噪声的互功率谱.改进后的加权函数为:

$$\psi_{xy} = \frac{1}{|G_{xy}(\omega) - G_{v_x v_y}(\omega)|} \quad (8)$$

其中, $G_{v_x v_y}(\omega)$ 表示 $x(n)$ 和 $y(n)$ 两传感器接收到的噪声的互功率谱.

### 3.3 信噪比加权

考虑到在实际情况中,整个频域内的信噪比不可能完全相同.也就是说,对于不同的 $\omega$ ,信噪比SNR的值也可能不同.为了加强互相关函数的锐化效果,得到更为精确的时延估计.对于信噪比较大的部分,应赋予一个更大的权值,而当信噪比较小时,赋予较小的权值.因此,构造一个新的加权函数:

$$\psi_{xy}(\omega) = \frac{1}{|G_{xy}(\omega) - G_{v_x v_y}(\omega)|^\rho}, \quad 0 \leq \rho \leq 1 \quad (9)$$

其中,加权因子 $\rho$ 随环境信噪比的变化而变化,其具体值需要多次通过实验测试优化得到.

### 3.4 相干函数的运用

在PHAT-GCC算法中,当信号能量较小时,加权函数的分母项趋于0,而加权函数整体趋于无穷大,从而会出现一个较大的误差.通常的做法是添加一个非零因子 $\beta$ 来解决该问题<sup>[10]</sup>.此时的加权函数为:

$$\psi_{xy}(\omega) = \frac{1}{|G_{xy}(\omega) - G_{v_x v_y}(\omega)|^\rho + \beta} \quad (10)$$

为了切合广义互相关算法的思想,本文考虑采用相干函数的概念作为非零因子 $\beta$ 的选择依据.对于文中两个麦克风阵元所接收到的语音信号,其功率谱的相干函数为:

$$\gamma_{xy}^2(\omega) = \frac{|G_{xy}(\omega)|^2}{G_x(\omega)G_y(\omega)}, \quad 0 < \gamma_{xy}^2(\omega) < 1 \quad (11)$$

在实际的系统中,两个麦克风阵元接收到的信号

不可能完全相干,也不可能完全无干,所以 $\gamma_{xy}^2(\omega)$ 介于0~1之间,这种情况也符合环境噪声对相干性的影响.用相干函数 $\gamma_{xy}^2(\omega)$ 来表征 $\beta$ ,保证了在信号能量较小时减小加权函数的误差,也避免了 $\beta$ 取值过大而过多地影响PHAT加权函数.

综上所述,最终改进的PHAT-GCC加权函数为:

$$\psi_{xy}(\omega) = \frac{1}{|G_{xy}(\omega) - G_{v_x v_y}(\omega)|^\rho + \gamma_{xy}^2(\omega)} \quad (12)$$

## 4 仿真结果与分析

为了比较分析不同加权函数的GCC时延估计和声源定位的性能,本文模拟麦克风阵元间距为 $d=0.1$  m的麦克风阵列接收语音信号.所采集的语音信号时长大约3 s,采样频率8 KHz,声源方向角 $\varphi=60^\circ$ .该语音信号的波形图如图4所示.

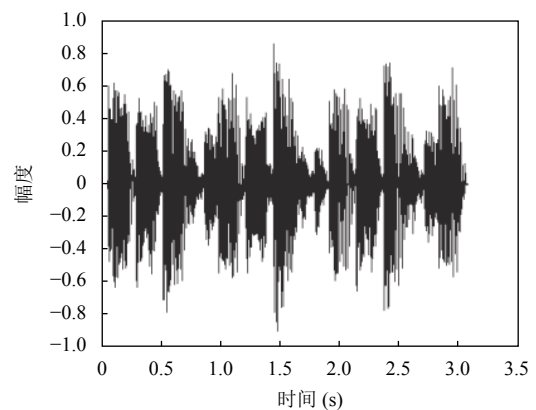


图4 语音信号

广义互相关函数的峰值尖锐程度是时延估计准确度的表征<sup>[2]</sup>,峰值越尖锐,时延估计越准确.在50 dB、20 dB和5 dB 3种信噪比情况下,选取CC、Roth、PHAT和改进的PHAT 4种加权函数做广义互相关MATLAB仿真.其中对于改进的PHAT-GCC算法,语音信号的预处理如图5所示.

对于4种加权函数的GCC算法仿真结果如图6~图8所示.

由图1麦克风阵列模型和式(1)可知,真正的时间延迟 $\tau$ 是一个关于入射角 $\varphi$ 的函数.因为 $\varphi$ 是从 $-90^\circ$ 到 $90^\circ$ 连续变化的物理量,这使得 $\tau$ 也是一个连续变量.而麦克风阵列中处理的是数字信号,即时间是一个个离散点.对于式(3)的参数 $T$ ,由于 $x(n)$ 和 $y(n)$ 是数字信号,所以 $T$ 只能取整数.因此, $T$ 实际上是时间延迟 $\tau$ 的一个近似值.

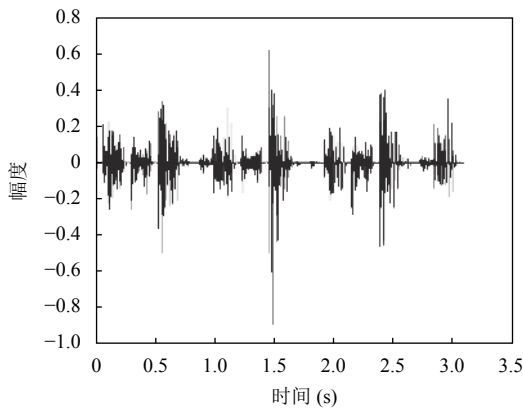


图5 改进的 PHAT-GCC 语音信号预处理

将仿真得到的 4 种加权函数的 GCC 时延估计代入式 (1), 比较在不同信噪比下, 声源定位的准确程度. 所得到的数据均保留两位小数, 结果如表 2 所示.

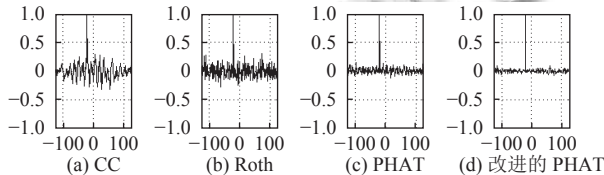


图6 信噪比 50 dB 时, 不同加权函数 GCC 的仿真结果

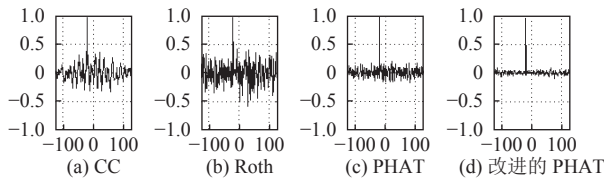


图7 信噪比 20 dB 时, 不同加权函数 GCC 的仿真结果

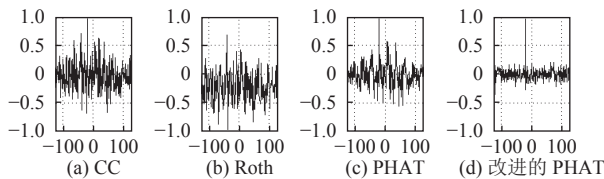


图8 信噪比 5 dB 时, 不同加权函数 GCC 的仿真结果

根据仿真结果, 在高信噪比情况下, 4 种加权函数得到的广义互相关函数峰值均较尖锐, 声源定位精度也很高. 但随着信噪比降低, 前 3 种加权方式的广义互相关函数的峰值尖锐程度降低, 同时干扰峰值随之增多, 有用信号相关函数的峰值也不够明显. 这表明 GCC 算法的时延估计准确性随信噪比的降低而恶化, 声源定位精度也随之骤降. 特别地, 对于 Roth 加权, 随着信噪比的降低, 性能急剧下降, 峰值受噪声影响严重, 几乎淹没在噪声中. 对于 PHAT 加权, 在高信噪比条件下, 峰值尖锐, 噪声波动较小, 被很好地抑制; 在低信噪比条件下, 相对于 CC 和 Roth 加权, 峰值更加尖锐, 保持着良好的抗噪特性. 此外, 相较于一般的 PHAT 加权, 改进后的 PHAT 加权法在低信噪比情况下性能明显更优, 峰值最为尖锐且噪声能够被更好地抑制掉, 声源定位的误差也是最小的. 综合来看, 改进的 PHAT 加权的广义互相关法表现出最优的时延估计性能, 而 Roth 加权的广义互相关法表现出最劣的时延估计性能.

### 5 结论与展望

本文通过 Matlab 仿真, 分析了 3 种常用加权函数的 GCC 时延估计算法的优劣性. 通过比较广义互相关函数、时延估计准确度和声源定位精度, 发现 PHAT 加权有更好的抗噪特性, 能够更加准确地做出时延估计. 但总体来说, GCC 时延估计的准确性随信噪比的降低, 性能逐渐下降. 因此, 本文提出了一种改进的 PHAT-GCC 算法. 从实验结果来看, 该算法在低信噪比情况下也能有效地抑制噪声尖锐峰值, 从而提高声源定位的精度. 这对麦克风阵列的研究和设计有着积极地指导意义. 在后续工作中, 为了进一步优化改进的 PHAT-GCC 算法, 考虑采用神经网络算法, 训练得到一个更优的参数组合, 从而提高该算法的性能.

表2 不同加权函数声源定位结果

信噪比 (dB)	CC			Roth			PHAT			改进的 PHAT		
	$\tau$ (ms)	$\varphi$ (°)	误差 (%)	$\tau$ (ms)	$\varphi$ (°)	误差 (%)	$\tau$ (ms)	$\varphi$ (°)	误差 (%)	$\tau$ (ms)	$\varphi$ (°)	误差 (%)
50	2.50	58.21	2.98	2.50	58.21	2.98	2.50	58.21	2.98	2.50	58.21	2.98
20	2.63	63.19	5.32	2.25	49.90	16.83	2.50	58.21	2.98	2.50	58.21	2.98
5	1.88	39.61	33.98	1.38	27.87	53.55	2.75	69.23	15.38	2.50	58.21	2.98

### 参考文献

1 Brandstein M, Ward D. Microphone Arrays: Signal Processing Techniques and Applications. Berlin: Springer-Verlag, 2001.

Verlag, 2001.

2 李幸远. 麦克风阵列语音增强系统设计与实现[硕士学位论文]. 西安: 西安电子科技大学, 2008.

- 3 Knapp C, Carter G. The generalized correlation method for estimation of time delay. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1976, 24(4): 320–327. [doi: [10.1109/TASSP.1976.1162830](https://doi.org/10.1109/TASSP.1976.1162830).]
- 4 Reed F, Feintuch P, Bershad N. Time delay estimation using the LMS adaptive filter-static behavior. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1981, 29(3): 561–571. [doi: [10.1109/TASSP.1981.1163614](https://doi.org/10.1109/TASSP.1981.1163614).]
- 5 Wiener N. Generalized harmonic analysis. *Acta Mathematica*, 1930, 55(1): 117–258.
- 6 Khintchine A. Korrelationstheorie der stationären stochastischen Prozesse. *Mathematische Annalen*, 1934, 109(1): 604–615. [doi: [10.1007/BF01449156](https://doi.org/10.1007/BF01449156).]
- 7 崔玮玮. 基于麦克风阵列的声源定位与语音增强方法研究 [博士学位论文]. 北京: 清华大学, 2009.
- 8 Omologo M, Svaizer P. Acoustic source location in noisy and reverberant environment using CSP analysis. *Proceedings of 1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*. Atlanta, GA, USA. 1996. 921–924. [doi: [10.1109/ICASSP.1996.543272](https://doi.org/10.1109/ICASSP.1996.543272).]
- 9 Quatieri TF. 离散时间语音信号处理. 赵胜辉, 译. 北京: 电子工业出版社, 2004.
- 10 李猛. 基于改进 GCC-PHAT 算法的麦克风阵列声源定位技术 [硕士学位论文]. 阜新: 辽宁工程技术大学, 2017.